

## **Supporting Information for: Understanding the errors of SHAPE-directed RNA structure modeling**

Wipapat Kladwang<sup>1</sup>, Christopher C. VanLang<sup>2</sup>, Pablo Cordero<sup>3</sup>, and Rhiju Das<sup>1,3,4\*</sup>

*Departments of Biochemistry<sup>1</sup>, Chemical Engineering<sup>2</sup>, Biomedical Informatics<sup>3</sup>, and Physics<sup>4</sup>, Stanford University, Stanford CA 94305*

\* To whom correspondence should be addressed: *rhiju@stanford.edu*. Phone: (650) 723-5976. Fax: (650) 723-6783.

This document contains the following Supporting Information:

### **Supporting Methods.**

**Table S1.** Benchmark for SHAPE-directed secondary structure modeling.

**Table S2.** Base-pair-level statistics of secondary structure recovery by *RNAstructure* with and without SHAPE data.

**Table S3.** Sources of poor discrimination of correct from incorrect secondary structures.

**Table S4.** Helix-by-helix bootstrap confidence estimates for the SHAPE-directed model of the HIV-1 RNA genome.

**Figure S1.** *RNAstructure* secondary structure models for a benchmark of six structured RNAs.

**Figure S2.** Recovery of SHAPE-directed model for a previously studied HCV RNA.

**Figure S3.** SHAPE data acquired with different dNTP mix for primer extension, refolding prior to chemical modification, and different DMSO backgrounds.

**Figure S4.** Demonstration that solution SHAPE data reflect folded or ligand-bound conformations.

**Figure S5.** Partition-function and bootstrap analysis of SHAPE-directed secondary structure models.

**Figure S6.** Sensitivity of minimum-energy model and robustness of bootstrap values to small changes in tRNA SHAPE data.

**Figure S7.** HIV-1 secondary structure helix confidence values compared to SHAPE reactivities.

**Figure S8.** Histogram and fit of SHAPE reactivities.

### **References for Supporting Information**

## Supporting Methods

*Likelihood-based attenuation correction, background subtraction, and normalization of SHAPE data*

Quantified SHAPE data were corrected for attenuation of longer reverse transcriptase products due to chemical modification, normalized, and background-subtracted. This procedure involved optimization carried out over three parameters, the overall modification rate  $\gamma$ , a normalization factor for the data  $\alpha$ , and a scaling for the background  $\beta$ , defined and estimated as follows. First, let  $y_i$  be the observed fraction of products stopping at each nucleotide  $i = 1, 2, \dots N$  in the reverse transcription reaction, ordered so that longer (attenuated) products have larger indices  $i$ ; and define

$y_{N+1} = 1 - \sum_{i=1}^N y_i$  be the fraction of products that are fully extended. These probabilities are

equal to the underlying stopping probabilities  $p_i$  times the product of probabilities that the reverse transcriptase has not stopped earlier:

$$y_i = p_i \prod_{j=1}^{i-1} (1 - p_j), \quad (1)$$

which can be inverted by recursively calculating  $p_1 = y_1$ , then  $p_2 = y_2(1-p_1)^{-1}$ , then  $p_3 = y_3(1-p_1)^{-1}(1-p_2)^{-1}$ , etc. For convenience, define  $c_i = p_i/y_i$  as an attenuation correction factor. [In the limit that the  $y_i$  are approximately constant and much smaller than 1, the solution reduces to an exponential fall-off  $c_i^{-1} = \exp[-\langle y_i \rangle i]$ , as was effectively assumed in (1, 2), but this approximation is unnecessary.]

Fluorescence measurements of reverse transcription products from capillary electrophoresis detectors are in arbitrary units; evaluating (1) requires that the  $y_i$  be

properly normalized, and this is in principle achieved by the constraint that

$$\sum_{i=1}^N y_i + y_{N+1} = 1.$$

However, in practice, the intensity of the longest products (e.g.,  $y_{N+1}$ )

cannot be measured precisely. The strong fluorescence of these fully extended products typically saturates the experimental detector. Thus, the values  $y_i$  are defined only up to a proportionality constant  $\gamma$ , i.e.,  $y_i = \gamma s_i / S$  where  $s_i$  are the observed fluorescence

intensities and  $S = \sum_{j=1}^{N_{\text{observed}}} s_j$ ; we need to select amongst  $\gamma < 1$ .

We estimate  $\gamma$  at the same time as the two other unknown proportionality constants in the data normalization and background subtraction. Let  $b_i$  be the quantified band intensities from control measurements (no SHAPE reaction), which we assume require negligible correction from attenuation. The background-subtracted SHAPE reactivities are given in terms of unknown constants  $\alpha$  and  $\beta$  by:

$$s_i^{\text{correct}} = \alpha[p_i - \beta b_i] = \alpha[c_i(\gamma)s_i - \beta b_i] \quad (2)$$

We then optimized the log-likelihood function:

$$L = \prod_{i=1}^n \alpha c_i(\gamma) \exp[F(\alpha[c_i(\gamma)s_i - \beta b_i])] \quad (3)$$

Here,  $F$  was chosen as a piecewise linear function that (i) gives an exponential distribution of positive SHAPE reactivities [similar to what is empirically observed (SI Fig. S8)], (ii) penalizes negative SHAPE reactivities, and (iii) results in a convex optimization problem for maximum likelihood estimation. The functional form was  $F(x) = F_+(x - x_0)$  if  $x > x_0$ , and  $F(x) = -F_-(x - x_0)$  if  $x \leq x_0$ . The parameters  $F_+ = 5.0$ ,  $F_- = 25.0$ , and  $x_0 = 0.06$  were used, and this formulation appears robust: varying  $F_+$  and  $F_-$  by

two-fold, changing  $x_0$  to zero, or using a double exponential fit for  $x > x_0$ , did not change the resulting corrected data beyond experimental errors (as estimated in the next section). We optimized (3) by performing a grid search of  $\zeta = \gamma/(1-\gamma)$  from 0.0 to 2.0 in 0.05 increments and iteratively solving for  $\alpha$  and then  $\beta$  (through relations obtained by  $dL/d\alpha = 0$  and  $dL/d\beta = 0$ ) until convergence. Applying the maximum likelihood values of  $\alpha$ ,  $\beta$ , and  $\gamma$  in (2) gave the corrected SHAPE reactivities. The algorithm is available in the function `overmod_and_background_correct_logL.m` within the freely available HiTRACE software package (3).

#### *Averaging across replicates, estimation of errors, and normalization*

The acquisition of multiple replicates across several experiments permitted high-quality final averaged data  $\bar{S}_i$  with error estimates  $\bar{\sigma}_i$ . To carry out the averaging, we noted that individual experiments might have different levels of measurement precision, and the variance of measurements within each experiment provide an estimate of that precision. These estimates, however, do not include systematic errors that differ between experiments, e.g., differing fluorescent backgrounds in different capillary electrophoresis instruments. We therefore carried out a two-part averaging. First, the data within each experiment  $j = 1, 2, \dots, M$  were averaged to give  $S_i^j$  and  $\sigma_i^j$ . As an example, suppose we have available 20 replicate measurements of each background/overmodification-corrected SHAPE profile  $s_i^k$ , where  $k = 1$  to 20. Suppose these data were measured for 4 independently prepared RNAs across  $M = 5$  different days/experiments. Then each  $k$  is the member of one and only one of 5 subsets  $E_j$ . Let  $N_j$  be the number of RNAs in set  $E_j$  (here  $N_j = 4$ ). For  $j = 1, 2, \dots, 5$ ,

$$S_i^j = \frac{1}{N_j} \sum_{k \in S_j} s_i^k \quad (4)$$

The estimated errors for these SHAPE data are:

$$\sigma_i^j = \left( \frac{1}{N_j} \sum_{k \in S_j} [s_i^k - S_i^j]^2 \right)^{1/2} \quad (5)$$

To combine measurements across multiple experiments, these merged data were averaged, with the inclusion of a position-dependent scale-factor  $\alpha_i$  that accounts for additional sources of experiment-to-experiment error. Explicitly, the assumed likelihood model was:

$$L(S_i) = \prod_j \frac{1}{2\pi\alpha_i\sigma_i^j} e^{-\frac{(S_i - S_i^j)^2}{2(\alpha_i\sigma_i^j)^2}} \quad (6)$$

This gives maximum-likelihood combined signal values  $\bar{S}_i$  and final Gaussian errors  $\bar{\sigma}_i$  of:

$$\begin{aligned} \bar{S}_i &= \frac{\sum_j [S_i^j / (\sigma_i^j)^2]}{\sum_j [1 / (\sigma_i^j)^2]} \\ \bar{\sigma}_i &= \alpha_i \left( \sum_j [1 / (\sigma_i^j)^2] \right)^{-1/2}. \end{aligned} \quad (7)$$

Here,  $\alpha_i$  is a scale factor and is again determined by optimizing the likelihood:

$$\alpha_i = \frac{1}{M} \sum_{j=1}^M \left[ (\bar{s}_i - s_i^j)^2 / (\sigma_i^j)^2 \right] \quad (8)$$

In practice, to obtain a robust estimate of this error scale factor, the average in (8) is taken across a 5-nucleotide window of bands around each nucleotide  $i$ . An example of this averaging is given in SI Fig. S3. These data, averaged across multiple replicates, were then normalized following a previously described procedure that was found to be optimal for *E. coli* ribosomal RNA (1). Briefly, the data sets were divided by a normalization factor, determined as the average of the top tenth percentile of band intensities. ‘Outliers’, identified as band intensities that exceeded 1.5 times the interquartile range, were removed before determining this factor. The resulting values lie mostly between 0 and 2 (see e.g., main text Fig. 1). The overall algorithm is available in the function `get_average_standard_state.m` within the freely available HiTRACE software package (3).

**Table S1. Benchmark for SHAPE-directed secondary structure modeling.**

RNA, source	Solution conditions <sup>a</sup>	Replicates <sup>b</sup>	Exp ts <sup>b</sup>	Off-set <sup>c</sup>	PDB <sup>d</sup>	Sequence & Secondary Structure <sup>e</sup>
tRNA <sup>phe</sup> , <i>E. coli</i>	Standard	14	7	-15	<b>1LOU</b> 1EHZ	gaaacaacaaaacaCGGAUUUAGCUAGGUUGGGAGAGGCCAGACUGAAGAUUCUG GAGGCCUGUUCGAUCCACAGAAUUCGCACCAaaaccaagaaacaacaacaac ac .....((((((.....))))..(((.....))) .....((((.....)))))))).... ...
P4-P6 domain, <i>Tetrahymena</i> ribozyme	Standard <sup>f</sup>	28	11	89	<b>1GTD</b> 1L8V 1HR2 2R8S	ggccaaaacaacgGAAUUGCAGGGAAAGGGGUCAACAGCCGUUCAGUACCAAGUCUA GGGGAAACUUUUGAGAUGGCCUUUCGAAGGGUAUUGGUAAAAGCUGACGGCAUGGU CUAACCACGCAGCCAAGGUCAACAGAUCUUCUGUUGAUUAGGAUGCAGUUC Aaaacccaaccaaaaacaacaacaacaac .....((((((.....))))..((.....))....)))).... .....))))((.....((((.....))))....))))....) .....
5S rRNA, <i>E. coli</i>	Standard	12	6	-20	<b>3OFC</b> 3OAS 3ORB 2WWQ ...	gaaaggaaaggaaagaaaUGCCUGGCCGUAGCGCGGUUGGUCCCCACUGACCC CAUGCCGAACUCAGAAGUGAAACGCCGUAGCGCCGAUGGUAGUGGGGUUCUCCCA UGCGAGAGUAGGGAACUGCCAGGCAUaaaacaaaacaagaacaacaacaac .....((((((.....((((.....))))....)))).... .....))))....))))....)))).... .....
Adenine riboswitch, <i>V. vulnificus</i> ( <i>add</i> )	Standard + 5 mM adenine	19	6	-8	<b>1Y26</b> 1Y27 299C 3GO2 ...	gaaaggaaaggaaagaaaCGCUCAUUAUACCUAAUAGUAUUGGUUUGGGAGUU CUACCAAGAGCCUUAAACUCUUGAUUAUGAAGUGaaaacaaaacaagaacaacaa caacaac .....((((((.....)))).... .....))))....))))....)))).... .....
c-di-GMP riboswitch, <i>V. cholerae</i> (VC1722)	Standard + 10 μM cyclic di-guanosine mono-phosphate	15	6	0	<b>3MXH</b> 3IWN 3MUV 3MUT ...	gaaaaauGUACGCACAGGGAAACAUUCGAAAGAGUGGGACGCAAAGCCUCGG CCUAAACAGAACAGCAUGGUAGGUAGCGGGGUUACCGAUGGCCAAUAGGcauacaaac caaagaacaacaacaac .....((((((.....)))).... .....))))....)))).... .....
Glycine riboswitch, <i>F. nucleatum</i>	Standard + 10 mM glycine	22	8	-10	<b>3P49</b>	ggacagagagGAUAUGAGGAGAGAUUCAUUUUAUGAAACACCGAAGAAGUAUC UUUCAGGUAAAAGGACUCAUUAUGGACGAACCUCUGGAGAGCUAUCUAAGAGUA ACACCGAAGGAGCAAGCUAAUUAUGGCCUAAACUCUCAGGUAAAAGGACGGAGaa acacaacaacaaacaacaacaac .....((((((.....)))).... .....))))....)))).... .....))))....)))).... .....

<sup>a</sup> Standard conditions are: 10 mM MgCl<sub>2</sub>, 50 mM Na-HEPES, pH 8.0 at 24 °C.

<sup>b</sup> All data average over experiments carried out on at least four different days to minimize systematic errors in sample preparations; within each experiment, two or more independently prepared and purified RNA stocks were assayed.

<sup>c</sup> Number added to sequence index to yield numbering used in previous biophysical studies, and in Figs. 1 and 2 of the main text.

<sup>d</sup> The first listed PDB ID was the source of the assumed crystallographic secondary structure; other listed IDs contain sequence variants, different complexes, or different crystallographic space groups and confirm this structure.

<sup>e</sup> In the sequence, lowercase symbols denote 5' and 3' buffer sequences, including primer binding site (last 20 nucleotides). In all cases, designs were checked in RNAsstructure and ViennaRNA to give negligible base pairing between added sequences and target domain. Structure is given in dot-bracket notation, and here denotes Watson/Crick base pairs for which there is crystallographic evidence. Only helices with two or more base pairs are included. For the adenine riboswitch, a two-base-pair helix [25-50, 26-49] that is not nested in the given secondary structure and involved in an extensive non-canonical loop-loop interaction is not included.

<sup>f</sup> Additional measurements were carried out with 30% methylpentanediol (MPD) due to reports that its presence in crystallization buffer can change SHAPE reactivity of the P4-P6 RNA (4). Measurements with MPD (10 replicates) gave different reactivities in the P5abc region; final SHAPE-directed secondary structure models mispredicted an additional helix compared to models guided by no-MPD data.

**Table S2. Base-pair-level statistics of secondary structure recovery by *RNAstructure* with and without SHAPE data.** TP=true positives; FP=false positives. TP' and FP' are the same, but allowing matches of base pair (i,j) with (i±1, j±1).

RNA	Len.	Number of base pairs									
		Cryst <sup>a</sup>	RNAstructure				+ SHAPE				
			TP	FP	TP'	FP'	TP	FP	TP'	FP'	
tRNA <sup>phe</sup>	76	20	12	12	12	12	15	6	15	4	
P4-P6 RNA	158	48	44	9	48	2	44	7	46	2	
5S rRNA	118	34	9	31	9	31	32	7	32	7	
Adenine ribosw.	71	21	15	10	15	10	21	2	21	2	
c-di-GMP ribosw.	80	25	21	5	21	3	21	6	21	4	
Glycine riboswitch	158	40	23	18	23	16	37	7	37	5	
Total	661	188	124	85	128	74	170	35	172	24	
<b>False negative rate<sup>b</sup></b>			34.0%		31.9%		9.6%		8.5%		
<b>False discovery rate<sup>c</sup></b>			40.7%		36.6%		17.1%		12.2%		
<b>Sensitivity<sup>d</sup></b>			66.0%		68.1%		90.4%		91.5%		
<b>Positive predictive value<sup>e</sup></b>			59.3%		63.4%		82.9%		87.8%		

<sup>a</sup> Cryst = number of helices in crystallographic model.

<sup>b</sup> False negative rate = 1 - TP/Cryst.

<sup>c</sup> False discovery rate = FP/(TP+FP).

<sup>d</sup> Sensitivity = (1 - false negative rate) = TP/Cryst.

<sup>e</sup> Positive predictive value = (1 - false discovery rate) = TP/(TP+FP).

**Table S3. Sources of poor discrimination of correct from incorrect secondary structures.**

Thermodynamic energies of base pairs and SHAPE pseudoenergies in kcal/mol, calculated in *RNAstructure*.

RNA	SHAPE-directed model			Crystallographic model <sup>a</sup>			Difference <sup>b</sup>		
	E <sub>total</sub>	E <sub>thermo</sub>	E <sub>SHAPE</sub>	E <sub>total</sub>	E <sub>thermo</sub>	E <sub>SHAPE</sub>	E <sub>total</sub>	E <sub>thermo</sub>	E <sub>SHAPE</sub>
tRNA <sup>phe</sup>	-40.1	-20.3	-19.8	-39.6	-20.5	-19.1	-0.5	0.2	-0.7
P4-P6	-125.6	-54.8	-70.8	-114.9	-46.4	-68.5	-10.7	-8.4	-2.3
5S rRNA	-95.5	-47.5	-48.0	-91.9	-45.7	-46.2	-3.6	-1.8	-1.8
Ade ribosw.	-48.2	-16.6	-31.6	-48.2	-16.6	-31.6	0.0	0.0	0.0
c-di-GMP ribosw.	-63.6	-26.3	-37.3	-62.7	-26.4	-36.3	-0.9	0.1	-1.0
Gly. ribosw.	-98.5	-24.8	-73.7	-96	-24.5	-71.5	-2.5	-0.3	-2.2
<i>Average</i>	-78.5	-31.7	-46.9	-75.5	-30.0	-45.5	-3.0	-1.7	-1.3

<sup>a</sup>For a fair comparison to the SHAPE-directed model, this is the lowest energy secondary structure produced by *RNAstructure* with the same SHAPE data, but forced to contain the crystallographically observed base pairs. For the adenine riboswitch, an ‘extra’ two-base-pair helix appears in this structure.

<sup>b</sup>Negative values indicate inaccuracy in structure discrimination.

<sup>c</sup>E<sub>total</sub> and E<sub>thermo</sub> are derived from from *efn2* (the *RNAstructure* package) run with and without SHAPE data, respectively. E<sub>SHAPE</sub> is the difference of the two values.

**Table S4. Helix-by-helix bootstrap confidence estimates for the SHAPE-directed model of the HIV-1 RNA genome.** Models were generated by applying RNAstructure 5.3 to SHAPE data from ref. (5). Following prior work, the temperature was set to the default (37 °C); slope  $m$  and intercept  $b$  of SHAPE pseudoenergy relation were set to 3.0 kcal/mol and -0.6 kcal/mol, respectively; maximum sequence distance between base pairs was set to 600; modeling was carried out for separate subsegments 1-2844, 2836-5722, 5676-6832, 6807-7791, and 7779-9173; and positions at termini of these subsegments (2836-2845, 5676-5724, 6799-6838, 7779-7791, 9171-9173) and in pseudoknotted regions (179-216, bound to tRNALys primer; 255-263 in the dimerization loop DIS; and 74-86 and 408-375, forming the 5' polyA signal) were forced to remain unpaired. “BP1”, “BP2”, and “len” give two residues marking the starting base pair of each stem and stem length; “P(boot)” and “BPP” are bootstrap confidence value and maximum Boltzmann probability in the stem as percentages; and “Modeled” gives whether the stem was in the working model of (5) and recovered with RNAstructure 5.3 (Y) or not (N), or whether the stem is newly predicted herein (X). Continued on next two pages.

BP1	BP2	len	P(boot)	BPP	Modeled	BP1	BP2	len	P(boot)	BPP	Modeled	BP1	BP2	len	P(boot)	BPP	Modeled
1	57	3	95.5	31.1	Y*	1214	1247	5	89	99.8	Y	2619	2666	4	96.5	100	Y
5	54	11	100	100	Y*	1223	1239	6	94	100	Y	2625	2661	4	91.5	98.9	Y
17	43	5	100	100	Y*	1249	1263	4	88.5	99.9	Y	2629	2654	10	100	100	Y
25	38	4	100	100	Y*	1350	1727	2	17	2.4	Y	2667	2686	3	47	6.5	Y
58	104	8	99.5	100	Y	1353	1724	5	20.5	2.8	Y	2671	2683	3	41	5.8	Y
67	94	3	99.5	100	Y	1360	1394	4	19.5	2	Y	2727	2745	4	89	100	Y
70	90	4	99.5	100	Y	1375	1387	5	69.5	95.3	Y	2731	2740	2	76	95.2	Y
106	343	9	40	98.5	Y	1396	1558	4	37	5.5	Y	2781	2802	7	92	100	Y
125	223	7	26.5	95.4	Y	1401	1554	3	38	5.6	Y	2811	2835	2	67	98.8	Y
134	178	8	77.5	99.4	Y	1405	1414	2	24	24.5	Y	2814	2833	6	80.5	100	Y
143	167	2	51.5	93.2	Y	1418	1457	3	29.5	21.8	Y	2846	3381	6	68.5	65.8	Y
146	164	2	64.5	94	Y	1421	1443	7	92.5	99.7	Y	2852	3374	5	53	64.6	Y
148	160	3	79	99.7	Y	1459	1522	6	54	29.4	Y	2876	3273	5	10.5	22.1	Y
228	334	6	11	27.1	Y	1465	1515	3	37.5	26.4	Y	2892	3176	5	16	22.4	Y
236	282	3	70	69.3	Y	1469	1511	3	34	26.2	Y	2908	3160	4	12.5	21.2	Y
243	277	4	93	77.2	Y	1477	1504	3	18.5	20.9	Y	2925	3129	3	7.5	17.5	Y
248	270	7	100	100	Y	1481	1500	4	14.5	15.6	Y	2939	3088	5	2.5	7.9	Y
283	299	3	37	70	Y	1531	1541	3	69.5	90.7	Y	2946	2953	2	22.5	18.1	Y
286	295	3	34.5	69.3	Y	1568	1707	10	100	100	Y <sup>b</sup>	2972	3038	5	8	8.5	Y
312	325	5	98	100	Y	1583	1694	3	53.5	68.3	Y <sup>b</sup>	2995	3006	4	38.5	46.6	Y
363	750	5	4.5	1.6	Y	1590	1683	6	48.5	77.8	Y <sup>b</sup>	3015	3023	3	28	45.9	Y
399	484	9	98	100	Y	1598	1640	2	35.5	65.8	Y <sup>b</sup>	3040	3058	3	36	56.1	Y
501	526	6	33	87	Y	1604	1636	8	78	98.7	Y <sup>b</sup>	3044	3055	2	35	55.9	Y
510	518	2	22	62	Y	1615	1625	3	97	100	Y <sup>b</sup>	3089	3105	2	24.5	17.7	Y
582	657	2	41.5	68.2	Y	1645	1672	12	100	100	Y <sup>b</sup>	3093	3101	3	41	38	Y
586	652	8	66.5	99.7	Y	1760	1785	4	27.5	28.4	Y	3140	3149	3	60.5	81	Y
595	625	4	58.5	91.6	Y	1767	1779	2	30	27.2	Y	3178	3190	3	40.5	91	Y
599	616	5	93	99.8	Y	1813	1916	6	53	94.9	Y	3205	3223	3	93	99.4	Y
628	636	3	36.5	83.8	Y	1823	1849	6	66	90.9	Y	3243	3256	4	50.5	47.1	Y
678	741	4	4	4.4	Y	1829	1842	3	88.5	99.4	Y	3285	3334	1	7	17.7	Y
683	691	2	25.5	12.1	Y	1862	1881	5	78	99.9	Y	3287	3332	6	15.5	22.7	Y
693	722	6	73.5	99.6	Y	1991	2326	9	57.5	100	Y	3297	3322	5	34	25.9	Y
702	714	2	59	93.4	Y	2015	2121	8	99.5	100	Y	3336	3358	5	8.5	25.4	Y
752	1172	5	22	36.9	Y	2024	2112	10	99	100	Y	3344	3351	2	9.5	22.3	Y
795	849	9	87.5	99.9	Y	2042	2070	7	82	99.9	Y	3393	3402	2	19	11.7	Y
821	831	4	88	98.2	Y	2051	2062	3	35	56.3	Y	3404	3943	4	11.5	5	Y
871	913	2	56	84.1	Y	2072	2082	3	25.5	23.3	Y	3410	3936	4	15.5	32.1	Y
874	911	6	77	99.3	Y	2135	2144	3	35.5	82.6	Y	3414	3931	7	18.5	33.8	Y
921	964	5	51	99.1	Y	2145	2171	6	58	91.6	Y	3423	3924	4	12	31.7	Y
926	935	3	20.5	68.1	Y	2201	2238	7	38.5	89.8	Y	3427	3496	4	16	37	Y
946	957	2	30	68.5	Y	2245	2260	4	35	91.5	Y	3432	3439	2	35	53.8	Y
1028	1064	5	67	82.2	Y	2268	2310	5	88	100	Y	3458	3468	3	44.5	89.4	Y
1076	1100	3	99	99.9	Y	2273	2301	5	88	99.8	Y	3497	3917	5	15	39.3	Y
1080	1097	5	100	100	Y	2279	2296	1	88	49.4	Y	3523	3653	6	15.5	5.6	Y
1102	1142	5	85	99.6	Y	2378	2429	8	68	99.2	Y	3532	3644	2	8	4	Y
1110	1137	3	84	99.9	Y	2391	2421	3	32	69.7	Y	3535	3550	3	31	53.5	Y
1116	1132	6	75	98.8	Y	2547	2778	9	52	0	Y	3581	3593	4	11	3.8	Y
1177	1312	5	64.5	39.2	Y	2558	2576	3	16	1.7	Y	3606	3639	2	19	41.4	Y
1183	1306	5	68.5	39.2	Y	2596	2712	5	31.5	0.4	Y	3609	3636	6	25	50.7	Y
1193	1299	6	83	99.2	Y	2603	2705	4	30.5	0.4	Y	3692	3907	6	18	61.6	Y
1214	1247	5	89	99.8	Y	2619	2666	4	96.5	100	Y	3699	3800	4	35	84.9	Y
1223	1239	6	94	100	Y	2625	2661	4	91.5	98.9	Y	3704	3759	7	94	99.9	Y
1249	1263	4	88.5	99.9	Y	2629	2654	10	100	100	Y	3818	3899	9	89.5	100	Y
1350	1727	2	17	2.4	Y	2667	2686	3	47	6.5	Y	3840	3890	6	91	100	Y
1353	1724	5	20.5	2.8	Y	2671	2683	3	41	5.8	Y	3857	3872	4	82.5	99.8	Y

<sup>a</sup> 5' TAR element.

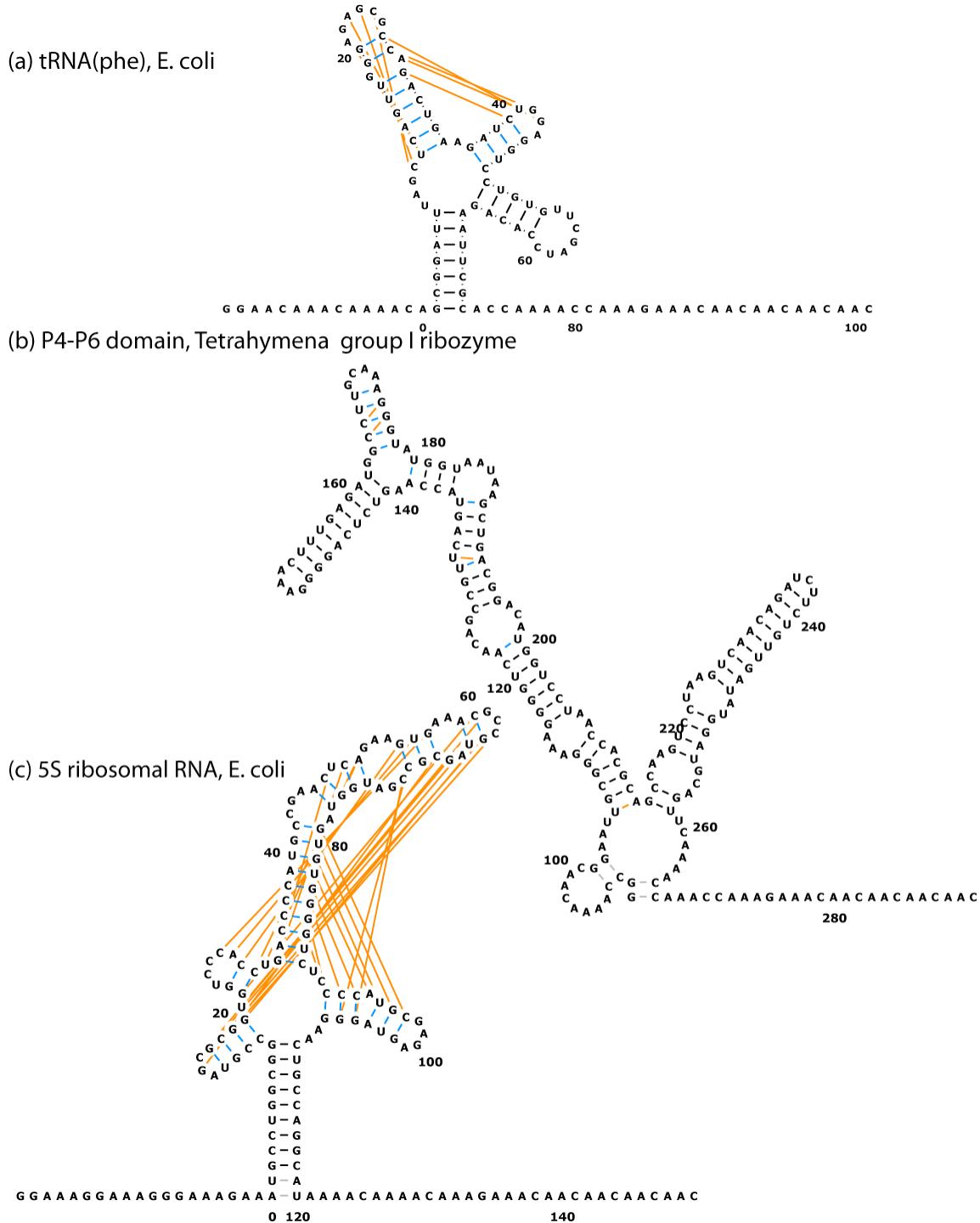
<sup>b</sup> gag-pol region.

BP1	BP2	len	P(boot)	BPP	Modeled	BP1	BP2	len	P(boot)	BPP	Modeled	BP1	BP2	len	P(boot)	BPP	Modeled
4204	4349	8	96.5	100	Y	5499	5519	3	47	79.8	Y	6538	6595	4	66	95	Y
4280	4299	4	91.5	96.8	Y	5503	5515	4	40	75.7	Y	6543	6590	3	45	33.4	Y
4301	4316	4	32.5	28.6	Y	5530	5581	5	24.5	69.1	Y	6546	6586	6	96	100	Y
4362	4503	7	45.5	52.4	Y	5536	5576	3	27	69.5	Y	6552	6579	4	90.5	99.7	Y
4447	4488	4	44	40.8	Y	5542	5571	2	31.5	70.5	Y	6618	6635	4	21.5	59.7	Y
4456	4479	2	34	31.5	Y	5545	5568	3	36	72.5	Y	6623	6630	2	18.5	56.4	Y
4458	4473	3	56	55.8	Y	5549	5565	2	15	49.1	Y	6706	6762	2	26	43.1	Y
4490	4496	2	32	36	Y	5600	5644	2	59.5	61.4	Y	6711	6757	9	90	100	Y
4551	5036	10	81	100	Y	5605	5639	6	95	100	Y	6839	7188	4	52	79.7	Y
4573	4586	4	26.5	53.3	Y	5612	5632	1	95	76.9	Y	6846	7179	2	71	97.1	Y
4588	4934	9	71	99.9	Y	5614	5631	4	95	99.9	Y	6850	7176	5	88	99.7	Y
4601	4914	8	87	100	Y	5725	6314	3	27.5	25.2	Y	6864	7113	5	10.5	16.8	Y
4614	4902	2	16	24.3	Y	5745	6243	10	97	99.9	Y	6870	6886	6	26	30.7	Y
4642	4674	3	28	85.4	Y	5763	6147	4	15.5	12.1	Y	6893	7077	1	19	7.9	Y
4646	4670	4	35	94.4	Y	5770	6142	6	37.5	16.9	Y	6894	7075	1	20	7.2	Y
4694	4732	3	31.5	86.2	Y	5793	6013	7	96.5	100	Y	6895	7073	7	97.5	100	Y
4698	4728	4	31.5	86.4	Y	5803	6004	5	91.5	99.9	Y	6904	7056	10	98	100	Y
4754	4770	6	100	100	Y	5816	5830	6	98.5	100	Y	6923	7040	4	78	97.9	Y
4797	4899	7	97.5	100	Y	5846	5861	6	98.5	100	Y <sup>c</sup>	6938	6956	5	81.5	96.8	Y
4807	4822	3	99.5	100	Y	5867	5997	7	94	100	Y	6964	6973	3	33.5	66.3	Y
4829	4891	9	100	100	Y	5874	5989	5	93.5	99.9	Y	6983	7016	3	13.5	33.9	Y
4840	4856	6	100	100	Y	5887	5894	2	26.5	63.6	Y	6988	7009	2	34.5	70.7	Y
4938	4999	6	81	100	Y	5896	5980	6	57.5	89.9	Y	6991	7007	4	40	80.5	Y
4951	4985	5	75.5	99.9	Y	5904	5913	2	36.5	63.6	Y	7079	7099	7	94	100	Y
4960	4980	6	99.5	100	Y	5915	5973	3	21	46.9	Y	7114	7136	7	94	98.9	Y
5010	5022	4	39.5	92.6	Y	5927	5962	4	24	41.4	Y	7150	7170	5	13	16.8	Y
5070	5100	6	19	47.2	Y	5931	5952	4	69	99.5	Y	7245	7599	1	66.5	80	Y
5083	5094	4	51	87.8	Y	5936	5947	4	78.5	100	Y	7247	7597	5	84	99.9	Y
5114	5132	4	73.5	4.8	Y	6024	6135	3	19.5	42.2	Y	7256	7590	8	99	100	Y
5139	5675	2	10	0.9	Y	6048	6066	5	67.5	92.4	Y	7272	7578	11	99.5	100	Y
5143	5673	7	20.5	1.7	Y	6072	6125	3	13	2.4	Y	7283	7566	3	99	100	Y
5154	5204	4	43	57.5	Y	6076	6121	2	12.5	2.4	Y	7291	7557	7	98.5	100	Y
5166	5194	5	75.5	85.2	Y	6078	6118	2	15.5	2.5	Y	7305	7538	7	94.5	100	Y
5206	5396	2	29	40.6	Y	6083	6113	3	22	4.5	Y	7312	7530	3	94	100	Y
5209	5394	5	77	100	Y	6092	6102	3	6.5	3.6	Y	7316	7526	5	94	100	Y
5216	5384	8	82	100	Y	6149	6159	2	2	5.4	Y	7321	7520	4	99.5	99.9	Y
5234	5265	3	29	56.9	Y	6185	6200	2	40.5	15.5	Y	7325	7515	4	99.5	100	Y
5239	5261	5	69	98.7	Y	6270	6290	6	69	98	Y	7333	7508	4	100	100	Y
5267	5297	6	65.5	97.9	Y	6328	6798	3	35.5	30.6	Y	7337	7503	5	100	100	Y
5273	5283	4	52	82.1	Y	6332	6375	6	20	34.8	Y	7343	7408	6	92.5	100	Y
5303	5343	4	50.5	96.6	Y	6381	6780	4	16	46.1	Y	7350	7378	1	67	95.5	Y
5307	5337	4	40	94.7	Y	6385	6775	3	15	47.7	Y	7353	7375	3	77	100	Y
5311	5331	6	76.5	98.1	Y	6393	6767	2	13	44.3	Y	7356	7371	2	82	100	Y
5349	5370	2	17	32.1	Y	6398	6704	5	30.5	88.6	Y	7358	7368	4	94.5	100	Y
5352	5367	1	17	20.1	Y	6416	6695	4	25	86.3	Y	7383	7399	6	100	100	Y
5355	5364	3	26	43	Y	6421	6521	8	26.5	85.8	Y	7411	7428	5	78	99.5	Y
5404	5419	5	99	100	Y	6432	6513	6	28	84.5	Y	7438	7464	3	63.5	98.1	Y
5425	5438	3	32.5	52.5	Y	6456	6467	4	41	88.4	Y	7443	7459	7	78	99.5	Y
5440	5450	3	28.5	40.4	Y	6475	6497	5	18	36.7	Y	7468	7493	9	100	100	Y
5473	5650	5	1	2.3	Y	6536	6598	2	39.5	67.5	Y	7601	7616	4	29	58.2	Y
5499	5519	3	47	79.8	Y	6538	6595	4	66	95	Y	7627	7636	2	48.5	58.4	Y
5503	5515	4	40	75.7	Y	6543	6590	3	45	33.4	Y	7647	7692	6	59.5	81.6	Y
5530	5581	5	24.5	69.1	Y	6546	6586	6	96	100	Y	7705	7770	6	67.5	99.8	Y
5536	5576	3	27	69.5	Y	6552	6579	4	90.5	99.7	Y	7712	7764	2	58	73.9	Y
5542	5571	2	31.5	70.5	Y	6618	6635	4	21.5	59.7	Y	7716	7760	2	61	96.4	Y

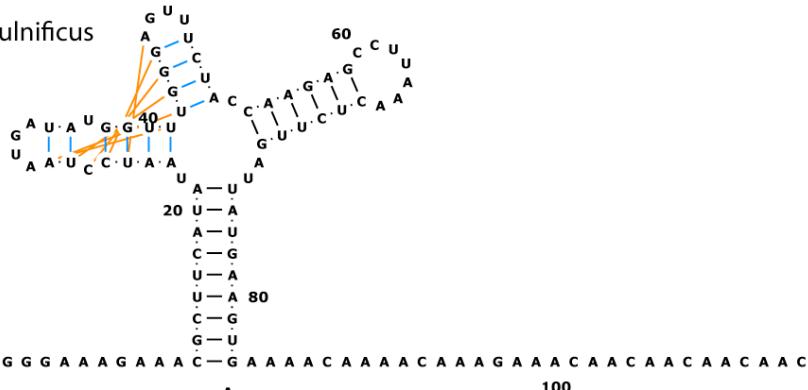
<sup>c</sup>signal-peptide stem at 5'end of gp120.

BP1	BP2	len	P(boot)	BPP	Modeled	BP1	BP2	len	P(boot)	BPP	Modeled	BP1	BP2	len	P(boot)	BPP	Modeled
7939	8107	2	13.5	76.6	Y	1341	1795	2	1.5	0.1	N	383	393	2	18.5	25.9	X
7941	8051	2	12	24.3	Y	1346	1790	4	1	0	N	380	396	2	14.5	17.6	X
7944	8049	5	50	97.8	Y	1729	1757	6	15	18.3	N	533	549	4	31	38.2	X
7949	8043	3	45	97.3	Y	1737	1749	3	26.5	21.4	N	972	979	2	17.5	4.6	X
7954	8039	3	19.5	31.9	Y	1796	1946	7	23.5	0.8	N	968	983	3	17.5	4.4	X
7969	8024	4	39	89.8	Y	1809	1920	2	20.5	2	N	966	984	2	8	1.7	X
7991	8015	5	94.5	100	Y	1811	1917	1	27.5	2.1	N	993	999	2	5	1.3	X
7997	8009	3	72	98.3	Y	1922	1930	3	29.5	18.2	N	987	1008	5	9	1.7	X
8053	8077	5	43	66.5	Y	1948	2545	7	7	0	N	770	1014	4	14	3.1	X
8059	8071	3	65.5	98.6	Y	2328	2348	3	17	1.7	N	765	1019	4	13.5	3	X
8083	8099	3	7.5	26.2	Y	2333	2343	3	17.5	1.6	N	761	1027	3	14.5	3.1	X
8169	8205	4	33.5	80.1	Y	2352	2520	8	7	0.5	N	1731	1752	5	40.5	74	X
8173	8194	3	38.5	64.4	Y	2363	2376	4	20.5	2.3	N	1795	1920	5	3.5	0.4	X
8309	8326	2	69	8.3	Y	2432	2484	6	30.5	1.1	N	1787	1925	3	3	0.3	X
8648	8667	3	45.5	66.1	Y	2448	2472	4	29.5	1.1	N	1345	1930	3	8.5	0	X
8651	8663	4	67.5	99.8	Y	2486	2497	2	2.5	0.2	N	2350	2357	2	19.5	35.8	X
8669	8679	3	98	100	Y	2714	2725	4	20	20.2	N	2348	2360	2	15	25.1	X
8686	9009	7	19	5.9	Y	3945	4518	4	7	0	N	2343	2366	5	17	47.2	X
8694	8999	4	12	5.7	Y	3958	3970	4	12.5	33.2	N	2430	2436	2	38	72.8	X
8700	8803	6	44.5	99	Y	4057	4068	2	31	24.5	N	1941	2453	13	51	0.1	X
8723	8751	11	96	100	Y	4539	5135	3	7.5	0	N	2496	2513	2	67.5	90.6	X
8753	8773	8	98.5	100	Y	7540	7546	2	22.5	43.4	N	2492	2517	3	65.5	92.3	X
8779	8790	3	34.5	81.8	Y	7638	7778	4	6.5	1.7	N	2477	2531	4	20	13.2	X
8807	8994	6	13	5.8	Y	7644	7774	3	3.5	0.6	N	2469	2544	8	37	17	X
8817	8985	5	31	6	Y	8226	8268	5	1	0	N	2721	2750	4	35	43	X
8830	8973	4	35	5.8	Y	8275	8348	5	26.5	3.8	N	2899	3167	3	5.5	9.8	X
8838	8965	2	21	4.7	Y	8282	8341	5	51	4.5	N	3765	3773	2	16.5	26.8	X
8851	8859	2	25	10.1	Y	8290	8308	3	19	1.7	N	3961	3970	2	5.5	19.6	X
8867	8906	10	99.5	100	Y	8358	8684	5	0.5	0.1	N	3958	4047	2	5.5	9	X
8878	8896	4	99	100	Y	8371	8412	1	5	1.1	N	3945	4058	4	11.5	30.8	X
8884	8890	1	98	91.6	Y	8373	8411	6	12.5	6.3	N	4131	4518	2	22	0.4	X
8912	8927	5	69.5	87.2	Y	8379	8404	1	11.5	4.1	N	6251	6258	2	24	65.6	X
9042	9057	5	84.5	99.7	Y	8382	8401	2	10.5	4.5	N	6390	6770	2	12.5	21.2	X
9059	9067	2	77	96.2	Y	8417	8641	7	3	77.2	N	7156	7164	2	13	13.4	X
9074	9134	5	98	99.9	Y	8432	8528	4	8	65.8	N	7621	7643	3	19	37.9	X
9080	9129	11	100	100	Y	8440	8519	3	9.5	58.9	N	7236	7777	4	47.5	95.9	X
9092	9118	5	100	100	Y	8455	8505	3	13	48.7	N	8265	8277	4	44.5	92.6	X
9100	9113	4	100	100	Y	8461	8498	11	28.5	62.6	N	8288	8339	6	39.5	1.1	X
9141	9170	4	76	70.6	Y	8530	8631	2	4	74.1	N	8282	8348	5	35.5	1.5	X
9145	9165	5	76	70.7	Y	8533	8628	10	16.5	89.3	N	8424	8435	3	14.5	17	X
382	537	4	6	40.5	N	8546	8617	4	8	54.9	N	8418	8442	4	28	12.7	X
547	565	4	37.5	58.5	N	8551	8565	3	6.5	60.4	N	8469	8478	2	14	6	X
760	1010	4	15	75.7	N	8578	8597	3	14.5	88.7	N	8465	8482	3	6.5	3	X
765	1005	2	15	75.5	N	8840	8935	2	1.5	0.2	N	8408	8493	6	10	5.7	X
769	1001	5	16	75.7	N	8845	8932	4	13	4.8	N	8406	8496	2	8.5	5.7	X
855	994	2	1.5	39.1	N	8949	8961	3	55	95.3	N	8404	8499	2	8.5	5.7	X
915	970	3	4	20.2	N	9011	9139	5	6.5	0	N	8398	8505	4	9	5.7	X
979	991	5	34.5	65.3	N	9020	9037	4	6.5	0	N	8392	8511	2	9	5.1	X
1026	1068	2	42	80	N	9025	9033	2	6.5	0	N	8370	8534	8	32	11	X
1200	1206	2	24.5	28.5	N	187	197	3	43	98.8	X	8367	8537	2	34.5	9.3	X
1341	1795	2	1.5	0.1	N	383	393	2	18.5	25.9	X	8358	8546	8	45	10.3	X
1346	1790	4	1	0	N	380	396	2	14.5	17.6	X	8228	8551	5	26	1.8	X
1729	1757	6	15	18.3	N	533	549	4	31	38.2	X	8224	8555	3	26	1.7	X
1737	1749	3	26.5	21.4	N	972	979	2	17.5	4.6	X	8596	8619	2	22	0.4	X
1796	1946	7	23.5	0.8	N	968	983	3	17.5	4.4	X	8591	8624	2	18	0.6	X

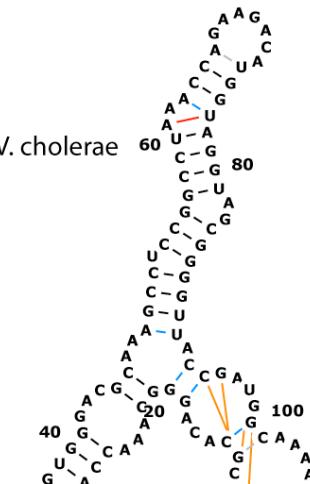
**Figure S1.** *RNAstructure* secondary structure models for a benchmark of six structured RNAs. Cyan lines mark incorrect base pairs; orange lines mark crystallographic base pairs missing in each model. (Figure is in two parts.)



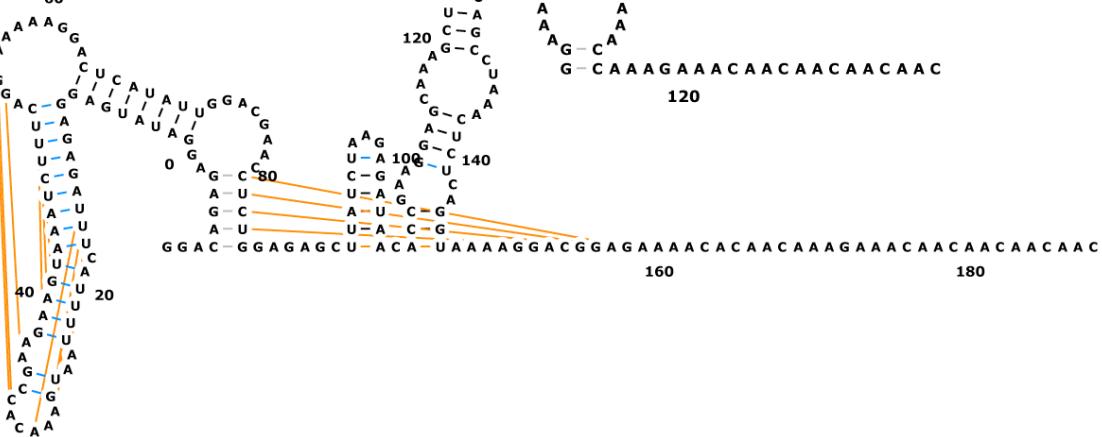
(d) adenosine riboswitch, *V. vulnificus*



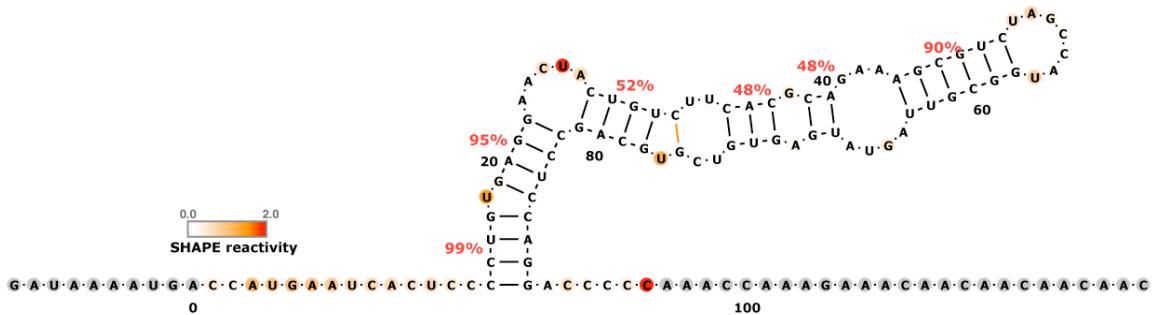
(e) cyclic diGMP riboswitch, *V. cholerae*



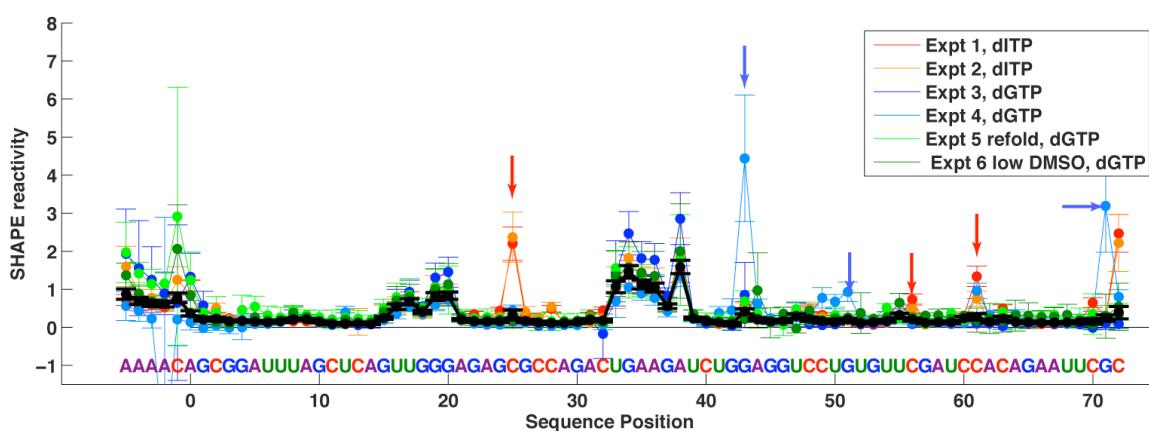
(f) glycine riboswitch, *F. nucleatum*



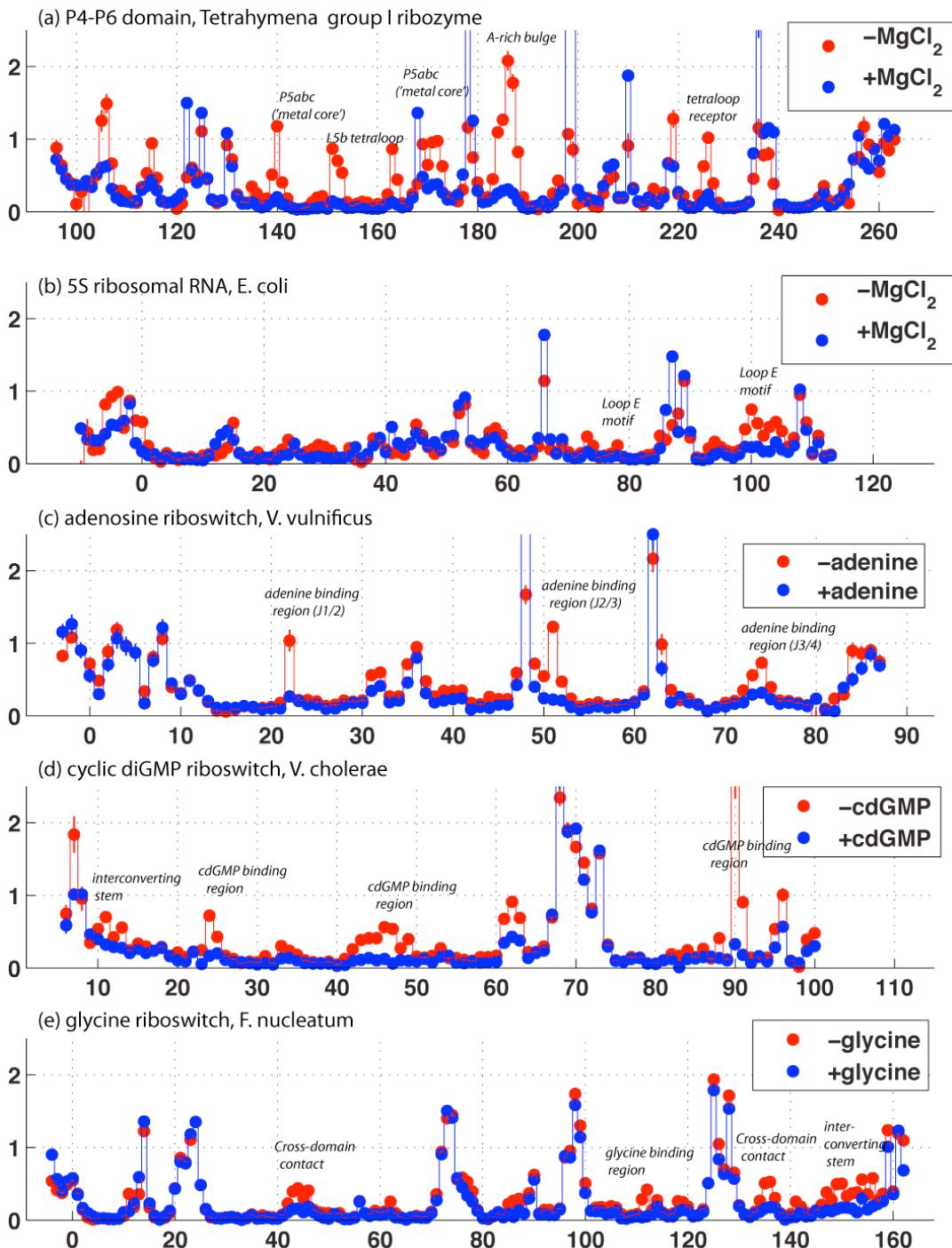
**Figure S2. Recovery of SHAPE-directed model for a previously studied HCV RNA.** As a control, SHAPE data were collected herein for the hepatitis C virus internal ribosomal entry site (HCV IRES) domain II; the resulting SHAPE-directed model agrees with prior work (1) and phylogenetic and NMR analyses [see, e.g., (6)]. Model is colored by SHAPE reactivity (see color scalebar). Helix confidence estimates from bootstrap analyses (see main text) are given as red percentage values. Flanking sequences (similar to those added to benchmark RNAs) are shown in gray.



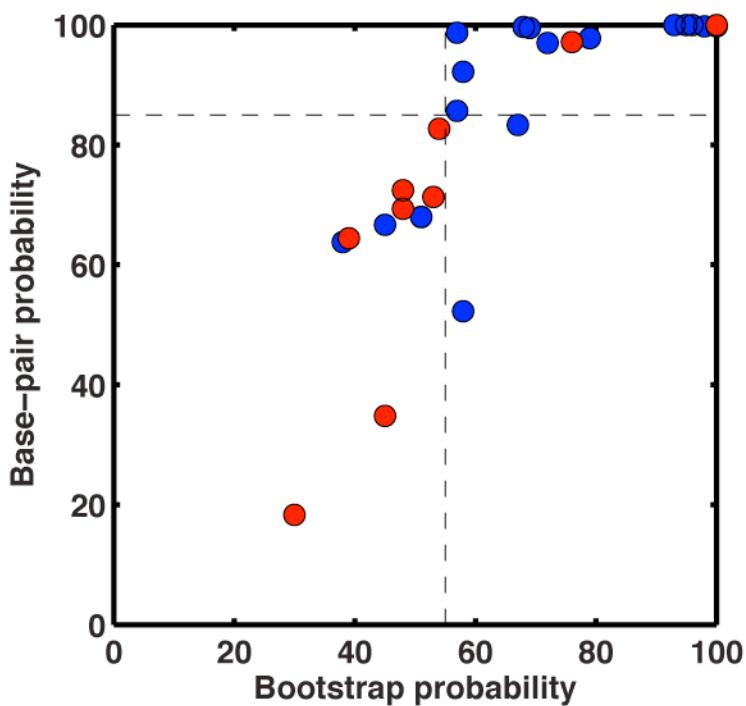
**Figure S3. SHAPE data acquired with different dNTP mix for primer extension, refolding prior to chemical modification, and different DMSO backgrounds.** Colored error bars and lines give background-subtracted data for tRNA<sup>phe</sup> (*E. coli*) from six experiments: two experiments in which the dATP, dCTP, dITP, and dTTP were used for reverse transcription of modified RNA; and four experiments in which standard dATP, dCTP, dGTP, and dTTP were used. Each experiment involved at least two replicate measurements; error bars represent standard deviations within each experiment. Arrows mark high-variance bands at C nucleotides in dITP experiments (red) due to poor incorporation of dITP, and near G nucleotides in dGTP experiments (blue) due to band compression. ‘Refold’ experiment 5 (green) involved incubation of RNA at 10 mM MgCl<sub>2</sub>, 10 mM Na-MES, pH 6.0 at 50 °C for thirty minutes and gave reactivities indistinguishable from conditions without incubation. Low DMSO experiment 6 (dark green) contained 10% DMSO during chemical modification and gave reactivities indistinguishable from conditions used for other experiments (25% DMSO). Black error bars and lines gives the final averaged SHAPE reactivity averaged over all experiments, taking into account the estimated errors (see SI Methods).



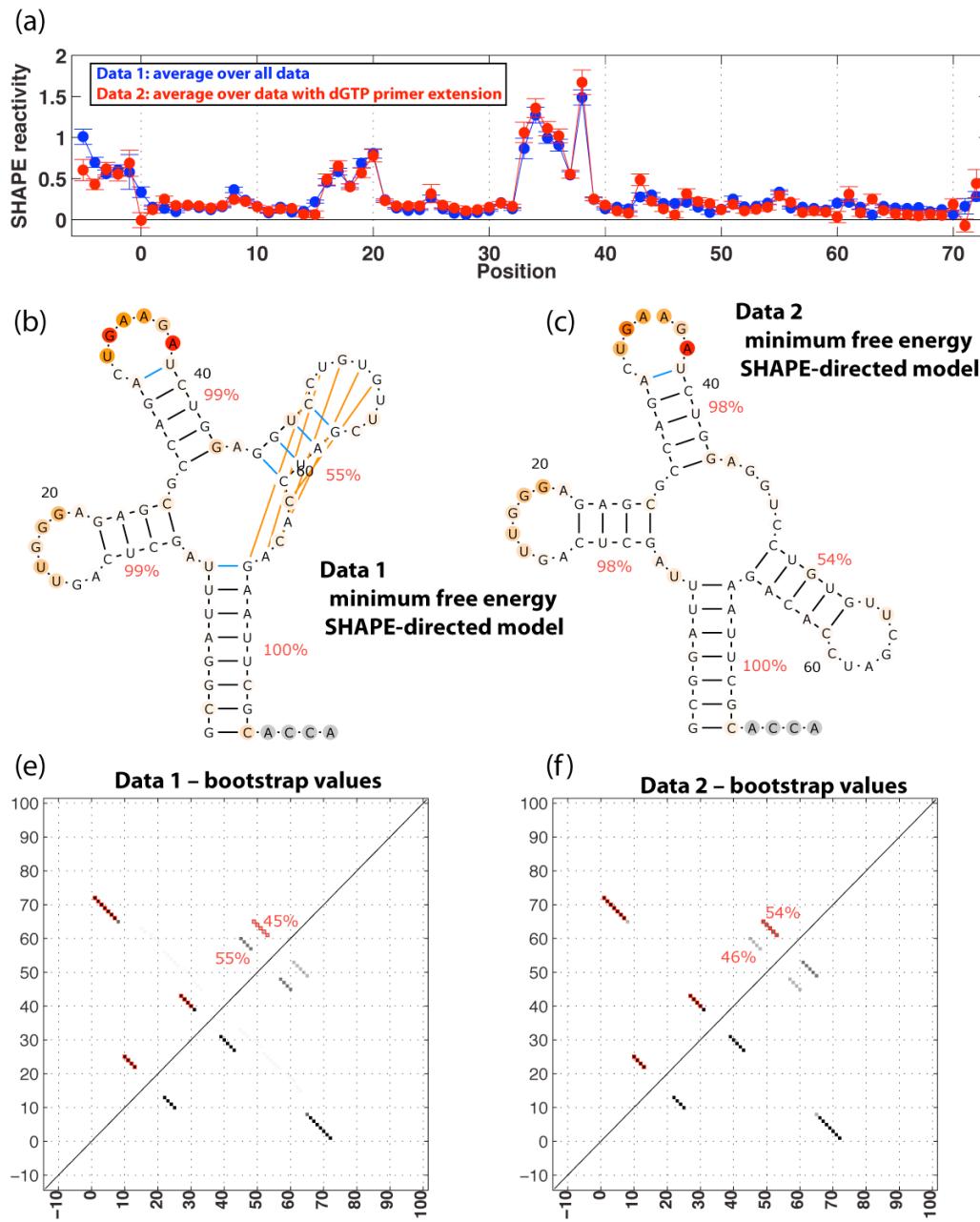
**Figure S4. Demonstration that solution SHAPE data reflect folded or ligand-bound conformations.** Significant differences were observed upon addition of 10 mM MgCl<sub>2</sub> with a background of 50 mM Na-HEPES (for the P4-P6 domain & 5S rRNA) and upon addition of ligand with a background of 10 mM MgCl<sub>2</sub>, 50 mM Na-HEPES, pH 8.0 (for the ligand-binding domains of riboswitches for adenine, c-di-GMP, and glycine). Regions that become protected upon Mg<sup>2+</sup>-induced tertiary folding or ligand binding are annotated on the data, and compare well to expectations from previous biophysical and crystallographic studies (7-13).



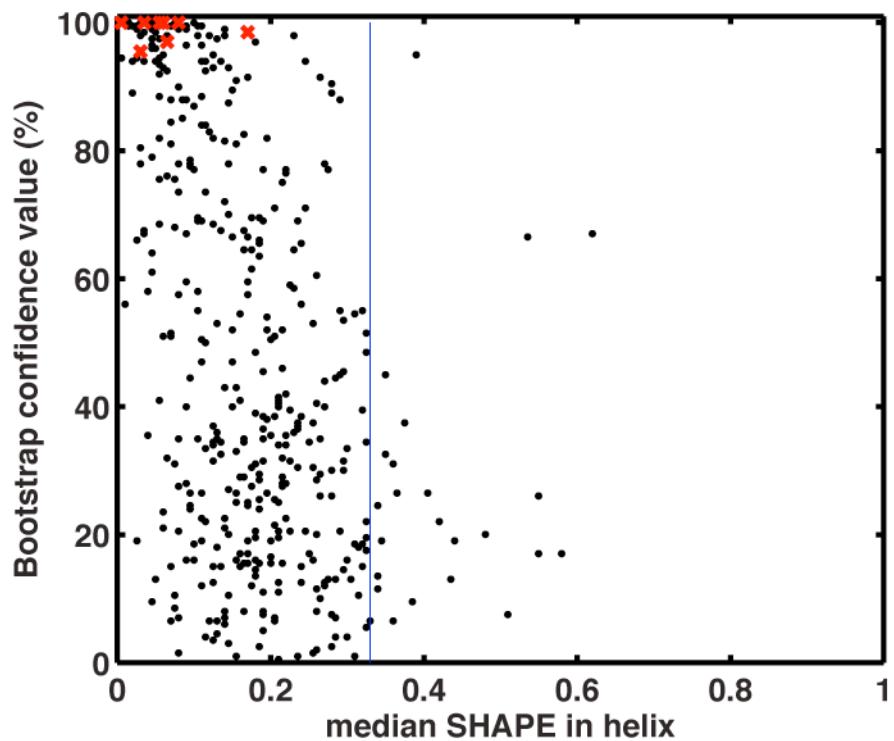
**Figure S5. Partition-function and bootstrap analysis of SHAPE-directed secondary structure models.** A confidence estimate for each helix in each of the six benchmark SHAPE-directed models was determined by (1) partition-function-based Boltzmann probabilities and (2) a nonparametric bootstrap analysis (repeating the modeling with ‘replicate’ data sets generated by randomly resampling the data with replacement). The confidence estimates for the two analyses correlate approximately, but partition function probabilities are skewed to higher values than bootstrap probabilities. Helices that agree (blue) or disagree (red) with crystallographic secondary structures are plotted separately. Dashed lines mark 80% and 55% separatrix values, above which two incorrect helices are observed, and 29 and 31 correct helices are observed for Boltzmann probability and bootstrap analyses, respectively.



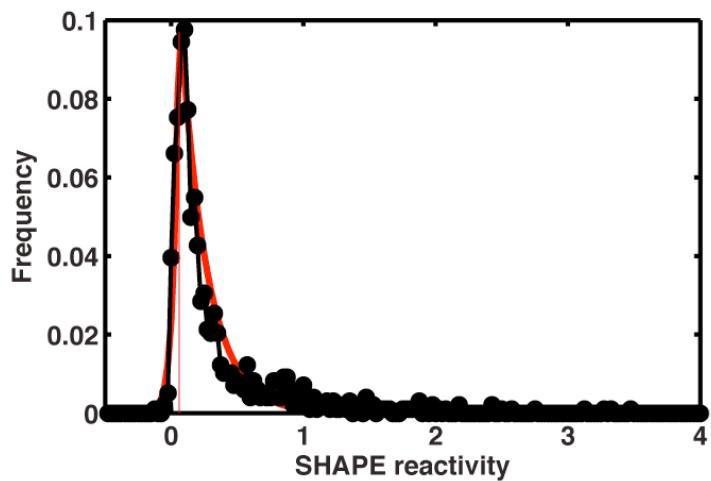
**Figure S6. Sensitivity of minimum-energy model and robustness of bootstrap values to small changes in tRNA SHAPE data.** (a) Comparison of SHAPE data sets obtained by averaging over all collected data (14 replicates; blue) and by averaging over just those data collected with primer extension with standard deoxynucleotide triphosphates (no dITP; i.e., dATP, dCTP, dGTP, dTTP) (6 replicates; red). (b,c) Minimum-energy SHAPE-directed secondary structures are different for the two data sets in the pairings of the third helix; bootstrap values given as red percentage values. (d,e) Helix probabilities from bootstrap analysis shown in grayscale, with 0 to 100% shown as white to black. Bootstrap values at alternate locations of third helix are shown as red percentage values; they are similar for the two data sets. Red squares mark crystallographic base pairs.



**Figure S7. HIV-1 secondary structure helix confidence values compared to SHAPE reactivities.** For each helix in the HIV-1 secondary structure model (5), the median SHAPE reactivity for nucleotides in the helix was computed, and plotted against bootstrap values. Blue line marks median reactivity over all nucleotides. High-bootstrap-value helices (e.g., four helices in TAR, three helices in gag-pol, and the gp120 signal-peptide stem; shown as red x's) typically have low median SHAPE reactivities. However, the converse is not true. Low-reactivity helices frequently have poor bootstrap values, indicating the existence of multiple secondary structures consistent with the data while still protecting the associated regions.



**Figure S8. Histogram and fit of SHAPE reactivities.** SHAPE reactivities of all residues for the six test RNAs (black; see SI Table S1), compared to least-squares fit (red) to a simple probability distribution  $P(x)$ . The distribution was assumed to take the form  $P(x) = N \exp(-F_+|x-x_0|)$  for  $x > x_0$ ; and  $P(x) = N \exp(F_-|x-x_0|)$  for  $x \leq x_0$ . The presented fit is for  $x_0 = 0.06$ ;  $F_+ = 5.0$ ;  $F_- = 25.0$ .



## References for Supporting Information

1. Deigan, K. E., Li, T. W., Mathews, D. H., and Weeks, K. M. (2009) Accurate SHAPE-directed RNA structure determination, *Proc Natl Acad Sci U S A* **106**, 97-102.
2. Vasa, S. M., Guex, N., Wilkinson, K. A., Weeks, K. M., and Giddings, M. C. (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis, *RNA* **14**, 1979-1990.
3. Yoon, S. R., Kim, J., and Das, R. (2011) HiTRACE: High Throughput Robust Analysis of Capillary Electropherograms, *Bioinformatics*, in press.
4. Vicens, Q., Gooding, A. R., Laederach, A., and Cech, T. R. (2007) Local RNA structural changes induced by crystallization are revealed by SHAPE, *Rna* **13**, 536-548.
5. Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr., Swanstrom, R., Burch, C. L., and Weeks, K. M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome, *Nature* **460**, 711-716.
6. Lukavsky, P. J. (2009) Structure and function of HCV IRES domains, *Virus Res* **139**, 166-171.
7. Takamoto, K., Das, R., He, Q., Doniach, S., Brenowitz, M., Herschlag, D., and Chance, M. R. (2004) Principles of RNA compaction: insights from the equilibrium folding pathway of the P4-P6 RNA domain in monovalent cations, *Journal of Molecular Biology* **343**, 1195-1206.
8. Correll, C. C., Freeborn, B., Moore, P. B., and Steitz, T. A. (1997) Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain, *Cell* **91**, 705-712.
9. Rieder, R., Lang, K., Gruber, D., and Micura, R. (2007) Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control, *Chembiochem* **8**, 896-902.
10. Mulhbacher, J., Brouillette, E., Allard, M., Fortier, L. C., Malouin, F., and Lafontaine, D. A. (2010) Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways, *PLoS Pathog* **6**, e1000865.
11. Smith, K. D., Lipchock, S. V., Livingston, A. L., Shanahan, C. A., and Strobel, S. A. (2010) Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch, *Biochemistry* **49**, 7351-7359.
12. Kulshina, N., Baird, N. J., and Ferre-D'Amare, A. R. (2009) Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch, *Nature structural & molecular biology* **16**, 1212-1217.
13. Kwon, M., and Strobel, S. A. (2008) Chemical basis of glycine riboswitch cooperativity, *RNA* **14**, 25-34.